THE DISTRIBUTION OF PRIME NUMBERS

VIVIAN KUPERBERG

CONTENTS

1.	Overview	1
2.	The Cramér Model	1
3.	Hardy–Littlewood k-tuples conjectures	5
References		8

1. OVERVIEW

The study of prime numbers is a beating heart of number theory. In this course we'll see analytic techniques applied to questions about prime numbers and thier distribution, and we'll get an understanding of what we should expect from prime numbers and from our proof techniques.

As an initial guide, here are some immediate questions one can (and does) ask about prime numbers:

- Are there infinitely many twin primes?
- How many twin primes are there?
- Are primes equidistributed mod q?
- How large does x have to be so that primes look equidistributed mod q? In other words, for which q can we show that primes up to x are equidistributed mod q?
- Can every even number ≥ 4 be written as a sum of two primes?
- Can very odd number ≥ 5 be written as a sum of three primes?
- What is the largest (and smallest) gap between primes $\leq x$?
- How often do large and small gaps occur?
- How many primes lie in the interval $[x, x + h(x)]$ for different functions $h(x)$? For which values of h is there an asymptotic answer? How much does this count vary?

The list goes on. Underpinning all of these questions is the following general question:

- In what ways do the primes behave like a "random sequence"? In what ways do they not?

We begin by formalizing what we mean by a "random sequence," through something called the Cramér model.

2. THE CRAMÉR MODEL

> "It is evident that the primes are randomly distributed but, unfortunately, we do not know what 'random' means."– R. C. Vaughan

The most basic quantitative result about prime numbers is the prime number theorem.

Theorem 2.1 (Prime Number Theorem, Hadamard–de la Vallée Poussin)**.** *Let $\pi(x)$ denote the number of primes $p \le x$. Then*

$$\pi(x) = \mathrm{li}(x)(1+o(1)) := \int_2^x \frac{\mathrm{d}t}{\ln t}(1+o(1)) = \frac{x}{\ln x}(1+o(1)).$$

Correspondingly, we would like a random model for the prime numbers that, at a minimum, agrees with the prime number theorem on average.

To begin with, we can discretize the logarithmic integral appearing in the prime number theorem into a sum over integers n:

Exercise 2.2. Show that

$$\int_2^x \frac{\mathrm{d}t}{\ln t} - \sum_{2\le n\le x} \frac{1}{\ln n} = O(\ln x).$$

The benefit of this expression is that, well, the integers are discrete. In particular, the statement "the prime number theorem should hold for a typical sequence in our random model of primes" is equivalent to saying "the probability that n is prime should be about $\frac{1}{\log n}$." Assuming that (and this is a big assumption!) each n is *independently* prime with probability $\frac{1}{\log n}$ leads us to the Cramér model for prime numbers.

Definition 2.3 (Cramér's random model)**.** For integers $n \ge 2$, let $X(n)$ be a sequence of independent random variables defined by

$$X(n) := \begin{cases} 1 & \text{with probability } \frac{1}{\log n} \\ 0 & \text{with probability } 1 - \frac{1}{\log n}, \end{cases}$$

which we think of as a randomized model for the behavior of $\mathbf{1}_{\mathcal{P}}$, the indicator function of the prime numbers.

By construction, we have

$$\mathbb{E} \sum_{2\le n\le x} X(n) = \sum_{2\le n\le x} \mathbb{E}X(n) = \sum_{2\le n\le x} \frac{1}{\ln n} = \mathrm{li}(x) + O(\ln x),$$

so we have correctly calibrated our random model to agree with PNT.

Now that we have the Cramér model, we can compare it to experimental data for various questions about prime numbers and see where we land.

Example 2.4 (Frequency of prime gaps)**.** A typical prime gap between p_k and the next prime is of size $\log p_k \sim \log k$, but one can ask a more refined question: for fixed $0 \le \alpha < \beta$ and primes p_k for $k \le n$, how often is $p_{k+1} - p_k \in [\alpha \log k, \beta \log k]$? That is, does the limit

$$\lim_{n\to\infty} \frac{1}{n} \# \left\{ 1 \le k \le n : \frac{p_{k+1} - p_k}{\log k} \in [\alpha, \beta] \right\}$$

exist, and what is it?

This probability is given by

$$\sum_{\alpha \log k \le h \le \beta \log k} \prod_{j=1}^{h-1} \left(1 - \frac{1}{\log(p_k + j)}\right) \frac{1}{\log(p_k + h)} \sim \sum_{\alpha \log k \le h \le \beta \log k} \left(1 - \frac{1}{\log k}\right)^{h-1} \frac{1}{\log k},$$

since $\log(p_k + j) \sim \log k$. This is

$$\sim \sum_{\alpha \le h/\log k \le \beta} e^{-h/\log k} \frac{1}{\log k} \sim \int_\alpha^\beta e^{-t} \mathrm{d}t.$$

So, the probability density of finding $p_{k+1} - p_k$ close to $t \log k$ is e^{-t}; in other words, primes are predicted to form a *Poisson process*.

Our next example will make use of the *Borel–Cantelli Lemma*:

Lemma 2.5 (Borel–Cantelli)**.** *Let $E_1, E_2, \dots$ be a sequence of events in a probability space. If the sum of the probabilities of the events $\{E_n\}$ is finite, i.e.*

$$\sum_{n=1}^{\infty} \mathbb{P}(E_n) < \infty,$$

then the probability that infinitely many events occur is 0*, that is,*

$$\mathbb{P}\left(\limsup_{n\to\infty} E_n\right) = 0.$$

Example 2.6 (Cramér and large gaps)**.** This example was in fact Cramér's motivation for defining the random model. Say we want to approximate the probability that the primes contain a gap of width at least g for some integer g. This is the same as saying that for some n, for all $1 \le v \le g$, each $n + v$ is nonprime (where we imagine that g grows with n). The probability of this event happening is

$$\prod_{v=1}^{g}\left(1 - \frac{1}{\log(n+v)}\right).$$

Say $g = c(\log n)^2$ for some constant $c > 0$. Then, discarding the small fluctuations in log, we get

$$\approx \left(1 - \frac{1}{\log n}\right)^{c(\log n)^2} \sim e^{-c\log n} = n^{-c}.$$

If $c > 1$, then the expected number of occurrences of gaps of size $c(\log n)^2$ is therefore

$$\sum_{n=2}^{\infty}\prod_{v=1}^{c(\log n)^2}\left(1 - \frac{1}{\log(n+v)}\right) \ll \sum_{n=2}^{\infty}\frac{1}{n^c} < \infty,$$

so we expect to see gaps of size $c(\log n)^2$ *only finitely often* by Borel–Cantelli (Lemma 2.5). By contrast if $c \le 1$, the expectation is

$$\sum_{n=2}^{\infty}\prod_{v=1}^{c(\log n)^2}\left(1 - \frac{1}{\log(n+v)}\right) \gg \sum_{n=2}^{\infty}\frac{1}{n^c} \to \infty,$$

so we expect to see gaps of width $c(\log n)^2$ infinitely often. Cramér conjectured therefore that

$$\limsup_{k\to\infty}\frac{p_{k+1} - p_k}{(\log p_k)^2} = 1, \tag{2.1}$$

where here p_k denotes the kth prime.

Let us record a few more predictions of the Cramér model before moving on to evaluating the strength of these predictions.

Example 2.7 (Cramér and Riemann)**.** What would the Cramér model say, for example, about the Riemann hypothesis? Recall that the *Riemann hypothesis*, in one of its many forms, is the statement that for any fixed $\varepsilon > 0$, we have

$$\sum_{n\le x}\Lambda(n) = x + O_\varepsilon(x^{1/2+\varepsilon}),$$

where $\Lambda(n)$ is the *von Mangoldt function* given by

$$\Lambda(n) = \begin{cases} \log p & \text{if } n = p^k \\ 0 & \text{else.} \end{cases}$$

We have immediately (this can be taken as an exercise) that

$$\sum_{\substack{p^k \le x \\ k \ge 2}} \Lambda(p^k) = O_\varepsilon(x^{1/2+\varepsilon}),$$

so we can restrict our attention to

$$\sum_{n \le x} \mathbf{1}_{\mathcal{P}}(n) \log n,$$

which we can model by

$$\mathbb{E} \sum_{2 \le n \le x} X(n) \log n.$$

By Borel-Cantelli (Lemma 2.5), it suffices to obtain the estimate

$$\sum_{2 \le n \le x} X(n) \log n = x + O_\varepsilon(x^{1/2} + \varepsilon) \tag{2.2}$$

with probability $1 - O(x^{-2})$, say, in which case exceptions happen only finitely often.

We consider the random variable $Y(n) := X(n) \log n - 1$; the $Y(n)$ have mean zero and are independent, and have size $O(\log n) = O(\log x) = O(x^{o(1)})$. Thus for any fixed natural number k,

$$\begin{aligned} \mathbb{E}\left(\sum_{2 \le n \le x} Y(n)\right)^k &= \mathbb{E} \sum_{2 \le n_1, \dots, n_k \le x} Y(n_1)Y(n_2)\cdots Y(n_k) \\ &= \sum_{2 \le n_1, \dots, n_k \le x} \mathbb{E}(Y(n_1)Y(n_2)\cdots Y(n_k)). \end{aligned}$$

Since the $Y(n)$ are independent with mean zero, any term $Y(n_1)Y(n_2)\cdots Y(n_k)$ where some n_i appears exactly once will have mean zero. Thus only terms with at most $k/2$ distinct indices n_i are nonzero; there are $O_k(x^{k/2})$ of these terms and each one contributes a power of $\log x = O(x^{o(1)})$, so the expectation overall satisfies

$$\mathbb{E}\left(\sum_{2 \le n \le x} Y(n)\right)^k = O(x^{k/2+o(1)}).$$

In particular the variance of $\sum_{2<n\le x} Y(n)$ is $O(x^{1+o(1)})$.

We now apply *Chebyshev's inequality*, which says that for a random variable X with variance σ^2 and mean 0, for any real number $r > 0$, for any natural number $k \ge 2$,

$$\Pr(|X| \ge r\mathbb{E}(|X|^k)^{1/k}) \le \frac{1}{r^k}.$$

In our case this tells us that

$$\Pr\left(\left|\sum_{2 \le n \le x} Y(n)\right| = O(x^{1/2+\varepsilon})\right) \ge 1 - O(x^{-k\varepsilon+o(1)}),$$

so taking k large enough depending on ε proves that

$$\sum_{2 \le n \le x} Y(n) = O_\varepsilon(x^{1/2+\varepsilon})$$

with probability $1 - O(x^{-2})$, and then expanding the definition of $Y(n)$ proves (2.2).

Note that in this example we really used very little about the Cramér model: just that the $Y(n)$ are bounded and that they have mean zero. This is, accordingly, a very general principle in probability: random variables with mean zero exhibit *square root cancellation*.

Finally, let's see what Cramér says about twin primes.

Example 2.8 (Cramér and twin primes)**.** According to the Cramér model, the number of twin primes $n \leq x$ is expected to be

$$\begin{aligned}\mathbb{E}\sum_{n\leq x} X(n)X(n+2) &= \sum_{n\leq x}\mathbb{E}(X(n)X(n+2))\\ &= \sum_{n\leq x}\mathbb{E}X(n)\mathbb{E}X(n+2)\\ &= \sum_{n\leq x}\frac{1}{(\ln n)(\ln(n+2))}\\ &\sim \sum_{n\leq x}\frac{1}{(\ln n)^2} \sim \mathrm{li}_2(x) \sim \frac{x}{(\ln x)^2},\end{aligned}$$

where the kth logarithmic integral $\mathrm{li}_k(x)$ is

$$\mathrm{li}_k(x) := \int_2^x \frac{\mathrm{d}t}{(\ln t)^k}.$$

This is not perfect but reasonably good if we compare to computer data! Computations suggest that the number of twin primes less than x is $\approx 1.32\mathrm{li}_2(x)$, so we seem to be correct to within a constant factor.

However, the example of twin primes suggests a different example where the Cramér model is hopelessly wrong. Say instead of twin primes we asked how many *consecutive* primes there are less than x, i.e., how many $n \leq x$ there are such that n and $n+1$ are both prime. We know the answer to be 1 for all $x \geq 2$, since $n = 2$ and $n+1$ is the only example; for larger n, either n or $n+1$ is even and thus composite. But the argument in Example 2.8 works in precisely the same way for consecutive primes as for twin primes, so the Cramér model predicts $\mathrm{li}_2(x)$ pairs of consecutive primes $\leq x$, and in particular infinitely many.

So, the primes are not actually independent: for example, if n is prime, then $n-1$ and $n+1$ *must* not be. However, the problem in the example of consecutive primes is specifically that we neglected to consider divisibility by 2, which suggests that we should adjust our model to include the influence of small prime numbers. This adjustment is best encapsulated by the *Hardy–Littlewood k-tuples conjectures*.

3. HARDY–LITTLEWOOD k-TUPLES CONJECTURES

The Hardy–Littlewood k-tuples conjectures, often referred to somewhat less descriptively as the "first Hardy–Littlewood conjecture," predict asymptotics for any linear correlations of prime numbers. Fix $k \geq 1$ and a k-tuple $\mathcal{H} = \{h_1, \ldots, h_k\}$ of *distinct* nonnegative integers. Then we may ask, as a generalization of our question about twin primes, how many values of $n \leq x$ have the property that $n + h_i$ is prime for every $1 \leq i \leq k$. In other words, if $\mathbf{1}_{\mathcal{P}}$ denotes the indicator of the set $\mathcal{P}$ of prime numbers, then

$$\sum_{n\leq x}\mathbf{1}_{\mathcal{P}}(n+h_1)\mathbf{1}_{\mathcal{P}}(n+h_2)\cdots\mathbf{1}_{\mathcal{P}}(n+h_k) = ?$$

Conjecture 3.1 (Hardy–Littlewood k-tuples conjecture)**.** *For a fixed integer $k \geq 1$ and a fixed k-tuple of distinct nonnegative integers $\mathcal{H} = \{h_1, \ldots, h_k\}$, as $x \to \infty$,*

$$\sum_{n\leq x}\prod_{i=1}^{k}\mathbf{1}_{\mathcal{P}}(n+h_i) = \mathfrak{S}(\mathcal{H})\mathrm{li}_k(x) + o(\mathrm{li}_k(x)), \tag{3.1}$$

where

$$\mathfrak{S}(\mathcal{H}) = \prod_{p \text{ prime}} \frac{1 - \nu_p(\mathcal{H})/p}{(1-1/p)^2},$$

and $\nu_p(\mathcal{H})$ *denotes the number of distinct congruence classes modulo* p *occupied by elements of* $\mathcal{H}$*.*

It will also be useful to us to consider the Hardy–Littlewood conjecture with von Mangoldt weights, given by the following:

Conjecture 3.2 (Hardy–Littlewood k-tuples conjecture with von Mangoldt weights)**.** *For a fixed integer* $k \geq 1$ *and a fixed k-tuple of distinct nonnegative integers* $\mathcal{H} = \{h_1, \dots, h_k\}$*, as* $x \to \infty$*,*

$$\sum_{n \leq x} \prod_{i=1}^{k} \Lambda(n + h_i) = \mathfrak{S}(\mathcal{H})x + o(x). \tag{3.2}$$

An enlightening interpretation of this conjecture is as a correction of the independence assumption that came from the Cramér model. The Cramér model would predict, just as we saw for twin primes, that the right-hand side of (3.1) should be $\mathrm{li}_k(x)$. The Cramér model, however, assumes that the primality of different integers is independent, which we can interpret as assuming that for every prime p and for distinct integers n and m, the event that n is divisible by p is independent from the event that m is divisible by p.

For a fixed prime p, the probability that k distinct randomly chosen integers are all nonzero mod p is $\left(\frac{p-1}{p}\right)^k = (1-1/p)^k$. But for a randomly chosen n, which we think of as large compared to both p and $h_1, \dots, h_k$, we have

$$n + h_i \not\equiv 0 \mod p \quad \forall i \Leftrightarrow n \not\equiv -h_i \mod p \quad \forall i.$$

This condition is precisely the condition that n does *not* land in one of $\nu_p(\mathcal{H})$ congruence classes mod p. Thus there are $p - \nu_p(\mathcal{H})$ congruence classes mod p for which n satisfies this condition, so the probability that a randomly chosen n has the property that $n + h_i \not\equiv 0 \mod p$ for all i is $\frac{p-\nu_p(\mathcal{H})}{p} = 1 - \frac{\nu_p(\mathcal{H})}{p}$. For each prime p, then, the pth component of $\mathfrak{S}(\mathcal{H})$ involves dividing by the independence assumption for divisibility by p and replacing it by the actual asymptotic probability that no element of the k-tuple is divisible by p.

Remark 3.3. Note that we are still using an independence assumption: we are assuming that the probability of being 0 mod $p-1$ and of being 0 mod p_2 for distinct primes p_1 and p_2 is independent. But this independence assumption has a strong justification: by Sunzi's remainder theorem, it is true! And not only do we get pairwise independence, but more generally: if q is any finite number, as $x \to \infty$, the Hardy–Littlewood conjectures (with the constant $\mathfrak{S}(\mathcal{H})$ restricted to be a product only over primes $p|q$) accurately predict the number of $n \leq x$ with $(n + h_i, q) = 1$ for all i.

Example 3.4 (Twin primes, revisited)**.** When $\mathcal{H} = \{0, 2\}$, the Hardy–Littlewood conjecture is precisely addressing the question of twin primes. In this case the prediction is that

$$\#\{p \leq x : p, p+2 \text{ prime}\} = 2 \prod_{p \geq 3} \frac{1 - 2/p}{(1-1/p)^2} \mathrm{li}_2(x)(1 + o(1)).$$

This prediction matches numerical data extremely well. The product over $p \geq 3$ (for reasons I don't understand, one usually excludes the 2) is called the *twin prime constant* and denoted C_2, so that $C_2 \approx 0.66016$ and the number of pairs of twin primes $\leq x$ is very close to 1.32032.

Example 3.5 (Consecutive primes, revisited)**.** At this point we have mostly exhausted the utility of this example, but let us revisit it once more to emphasize that it fails for the "correct" reason.

Here we would like to predict that there are only finitely many consecutive primes, so when $\mathcal{H} = \{0, 1\}$ we should have $\mathfrak{S}(\mathcal{H}) = 0$. Indeed, for $p = 2$, $v_p(\{0, 1\}) = 2$, since 0 and 1 take up two congruence classes mod 2. Thus $\frac{1-v_2(\mathcal{H})/2}{(1-1/2)^2} = \frac{1-2/2}{(1-1/2)^2} = 0$, so $\mathfrak{S}(\mathcal{H}) = 0$. This is capturing precisely the same reason we discussed before: the fact that $v_2(\{0,1\}) = 2$ *is* the fact that for all n, either n or $n+1$ is even, or 0 mod 2.

By the same argument, one can show that whenever $\mathfrak{S}(\mathcal{H}) = 0$, there are only finitely many n such that $n + h_i$ is prime for all i. A tuple $\mathcal{H}$ is called *admissible* if $\mathfrak{S}(\mathcal{H}) \neq 0$.

3.1. **Hardy–Littlewood and primes in short intervals.** The power, in some cases, of the Cramér model, is that beginning with the Hardy–Littlewood conjecture instead, one sometimes gets the same answer. In this section, we'll outline an argument due to Gallagher [2] about primes in intervals of length $\lambda \log x$, for some real constant $\lambda > 0$. Much of this exposition follows [4].

Let $h = \lambda \log x$. To understand $\pi(n+h) - \pi(n)$ as n varies over integers $n \leq x$, consider the rth moment

$$\frac{1}{x}\sum_{n\leq x}(\pi(n+h)-\pi(n))^r = \frac{1}{x}\sum_{n\leq x}\Big(\sum_{\substack{\ell=1 \\ n+\ell \text{ prime}}}^{h} 1\Big)^r. \tag{3.3}$$

The Cramér prediction would say that the rth moment is approximately

$$\frac{1}{x}\mathbb{E}\Big(\sum_{2\leq n\leq x}\Big(\sum_{\ell=1}^{h} X(n+\ell)\Big)^r\Big). \tag{3.4}$$

If (3.3) and (3.4) are roughly equal for $r \leq R$, as long as $R = R(x)$ tends to infinity, then $\pi(n+h) - \pi(n)$, for n chosen uniformly at random in $[1, x]$, has a Poisson distribution with parameter λ, because the Poisson distribution is determined by its moments. (See exercises for a bit more on this).

Expanding the rth power in (3.3), we get

$$\sum_{1\leq \ell_1,\ldots,\ell_r\leq h} \mathbf{1}_{\mathcal{P}}(n+\ell_1)\cdots\mathbf{1}_{\mathcal{P}}(n+\ell_r).$$

This is something we'd like to apply Hardy–Littlewood to, but we note that the ℓ_i might not all be different, so we have to do a little counting in order to account for this. Suppose there are exactly k distinct numbers among the $\ell_1, \ldots, \ell_r$, and say that they are given by $1 \leq h_1 < h_2 < \cdots < h_k \leq h$. The number of choices for $\ell_1, \ldots, \ell_r$ that yield the same distinct, ordered list $h_1, \ldots, h_k$ is given by the number of different ways of mapping $\{1, 2, \ldots, r\}$ onto $\{1, \ldots, k\}$. This number is called a *Stirling number of the second kind*, and we will denote it by $\sigma(r,k)$. Then (3.3) can be written as

$$\sum_{k=1}^{r}\sigma(r,k)\sum_{1\leq h_1<h_2<\cdots<h_k\leq h}\left(\frac{1}{x}\sum_{n\leq x}\mathbf{1}_{\mathcal{P}}(n+h_1)\cdots\mathbf{1}_{\mathcal{P}}(n+h_k)\right)$$
$$\sim \sum_{k=1}^{r}\frac{\sigma(r,k)}{(\log x)^k}\sum_{1\leq h_1<\cdots<h_k\leq h}\mathfrak{S}(\{h_1,\ldots,h_k\}),$$

where the right-hand side follows from the left by Hardy–Littlewood. Applying the same logic to the Cramér model expression in (3.4), we get that the two models agree if and only if

$$\sum_{1\leq h_1<\cdots<h_k\leq h}\mathfrak{S}(\{h_1,\ldots,h_k\}) = \sum_{1\leq h_1<\cdots<h_k\leq h} 1, \tag{3.5}$$

that is, if the singular series constants $\mathfrak{S}(\{h_1, \ldots, h_k\})$ are 1 on average.

Proposition 3.6 (Gallagher, 1976 [2])**.** *For each $k \geq 1$, the singular series is* 1 *on average for sets of size k, that is,* (3.5) *holds for fixed k as h grows large.*

Proof sketch. The first step is to prove that, for any fixed prime p,

$$\sum_{1\le h_1<\cdots<h_k\le h} \frac{(1-\nu_p(\mathcal{H})/p)}{(1-1/p)^k} \sim \sum_{1\le h_1<\cdots<h_k\le h} 1,$$

with a certain error term.

This is the same as saying that if we define

$$1+a_p(\mathcal{H}) = \frac{1-\nu_p(\mathcal{H})/p}{(1-1/p)^k},$$

then $a_p(\mathcal{H})$ converge to 0 on average as h grows large. We can extend $a_p(\mathcal{H})$ to any squarefree q, defining $a_q(\mathcal{H}) := \prod_{p|q} a_p(\mathcal{H})$ and $a_1(\mathcal{H}) = 1$, so that

$$\mathfrak{S}(\mathcal{H}) = \sum_{q\ge 1} a_q(\mathcal{H}).$$

By Sunzi's remainder theorem (although this is by no means immediate), $a_q(\mathcal{H})$ also converges to 0 on average over $\mathcal{H}$ as h grows large, for any $q > 1$.

Then the sum we want to estimate is

$$\sum_{1\le h_1<\cdots<h_k\le h} \mathfrak{S}(\mathcal{H}) = \sum_{1\le h_1<\cdots<h_k\le h} \sum_{q\ge 1} a_q(\mathcal{H}).$$

When q is "small" in some sense with respect to h, we can use that $a_q(\mathcal{H})$ converges to 0 on average over $\mathcal{H} \subseteq [1, h]$. When q is large, we can instead use that the tail of the singular series $\sum_{q>y} a_q(\mathcal{H})$ is small. □

This tells us that, if (and since) we believe the Hardy–Littlewood conjectures, we should expect the Cramér guess to be correct for the number of primes in log-size intervals: the distribution should indeed be Poissonian.

The Cramér model is not always so lucky. For example, in [3] it is shown that the sum of the von Mangoldt function in intervals $(n, n+H]$ where $1 \le n \le x$ and $x^\delta \le H \le x^{1-\delta}$, for a constant $\delta > 0$, has a distribution that is Gaussian (as one would expect from Cramér) with mean $\sim H$ and variance $\sim H \log \frac{N}{H}$ (which is smaller than the variance $\sim H \log N$ that Cramér would imply). So there is a balance to be explored: the Cramér model is a useful back-of-the-envelope calculation to know approximately what we should expect from questions about prime numbers, but it does not always agree with the inside-of-the-envelope calculation coming from other more sophisticated predictions which we believe much more.

REFERENCES

1. K. Ford, *Sieve methods lecture notes, spring 2023*.
2. P. X. Gallagher, *On the distribution of primes in short intervals*, Mathematika **23** (1976), no. 1, 4–9. MR 409385
3. H. L. Montgomery and K. Soundararajan, *Primes in short intervals*, Comm. Math. Phys. **252** (2004), no. 1-3, 589–617. MR 2104891
4. K. Soundararajan, *The distribution of prime numbers*, Equidistribution in number theory, an introduction, NATO Sci. Ser. II Math. Phys. Chem., vol. 237, Springer, Dordrecht, 2007, pp. 59–83. MR 2290494